TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.05, I.02 (2026)
https://www.tejasjournals.com/
https://doi.org/10.63920/tjths.52031

Fake Job Posting Detection Using LSTM-Based Deep Learning Model

Khushi Yadav[a], Prince Kannaujiya[b], Annapurna Pandey[c], Akarsh Yadav[d]

[a,b,c]Scholar, Department of Computer Science and Engineering, (AI&ML), KIPM College of Engineering and Technology, U.P., India
d Assistant Professor, Department of Computer Science and Engineering, (AI&ML) , KIPM College of Engineering and Technology, U.P., India
khushiyadav88400@gmail.com, princekannaujiya52@gmail.com, ap4806372@gmail.com, akarsh.9565@gmail.com

KEYWORD

Fake Job Posting Detection, LSTM, NLP, Deep Learning, Fraud Detection, Text Classification

ABSTRACT

With the rapid growth of online recruitment platforms, fake job postings have become a major concern, leading to financial loss and misuse of personal information of job seekers. This paper presents an intelligent approach for detecting fraudulent job postings using a Long Short-Term Memory (LSTM) based deep learning model combined with Natural Language Processing (NLP) techniques. The proposed system analyzes textual features such as job descriptions, requirements, and company profiles to classify postings as genuine or fake. Various preprocessing techniques including tokenization, stopword removal, and text normalization are applied to enhance the quality of input data. The LSTM model effectively captures sequential dependencies in textual data, improving the model's ability to detect hidden patterns in fraudulent content. The performance of the model is evaluated using standard metrics such as accuracy and precision, showing promising results. The proposed approach can assist job portals and users in identifying suspicious listings and reducing the risk of online recruitment fraud.

1. Introduction

In recent years, online recruitment platforms have gained significant popularity as they provide a convenient and efficient way for job seekers and employers to connect. However, along with their rapid growth, there has been a noticeable increase in fraudulent job postings. These fake job listings are designed to mislead candidates, often resulting in financial loss, identity theft, and misuse of sensitive personal information. The presence of such fraudulent activities has created a serious challenge for both users and online job portals.

Detecting fake job postings is a complex task because many fraudulent listings closely resemble genuine ones in terms of structure, language, and content. Traditional detection methods, such as manual verification and rule-based filtering systems, are often insufficient due to the large volume of data and the evolving nature of fraud techniques. As a result, there is a growing need for automated and intelligent systems that can effectively identify and filter out fake job postings.

With the advancement of Artificial Intelligence (AI) and Machine Learning (ML), more sophisticated approaches have been developed to address this problem. In particular, Natural Language Processing (NLP) techniques have shown promising results in analyzing textual data and extracting meaningful patterns from

Corresponding Author: Khushi Yadav, Scholar, Department of Computer Science and Engineering, (AI&ML), KIPM College of Engineering and Technology, U.P., India
Email: khushiyadav88400@gmail.com

TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.05, I.02 (2026)
https://www.tejasjournals.com/
https://doi.org/10.63920/tjths.52031

job descriptions and requirements. Deep learning models, especially Long Short-Term Memory (LSTM) networks, are highly effective for processing sequential data and understanding contextual relationships within text.

This paper proposes an intelligent system for detecting fraudulent job postings using an LSTM-based deep learning approach combined with NLP techniques. The model processes textual features from job listings and classifies them as genuine or fake based on learned patterns. The objective of this research is to improve the accuracy and reliability of fake job detection systems and contribute to creating a safer online recruitment environment for users.

2. Literature Review

The problem of detecting fraudulent job postings has attracted significant attention in recent years due to the increasing number of online recruitment scams. Various researchers have proposed different machine learning and natural language processing techniques to address this issue. Traditional approaches primarily relied on supervised machine learning algorithms such as Naive Bayes, Support Vector Machines (SVM), and Decision Trees to classify job postings as real or fake based on textual and categorical features.

Table 1. Data characteristics.

Sl	Feature	Description
1	job_id	A unique identifier for each job posting.
2	location	The geographical location of the job.
3	department	The department or organizational unit of the job belongs.
4	salary_range	The salary range for the job.
5	company_profile	A brief description of the company.
6	description	The detailed job description.
7	requirements	A list of required skills or qualifications for the job.
8	benefits	The benefits offered by the company.
9	telecommuting	A binary variable indicating whether the job allows telecommuting.
10	has_company_logo	A binary variable indicating a company logo's presence in the job posting.
11	has_questions	A binary variable indicating whether the job posting includes screening questions.
12	employment_type	The type of employment.
13	required_experience	The required experience level for the job.
14	industry	The industry to which the job belongs.
15	function	The job function or role.
16	fraudulent	Target variable indicating whether the job is genuine (0) or fake (1).

Several studies have utilized Natural Language Processing (NLP) techniques to extract meaningful information from job descriptions, including keyword extraction, term frequency-inverse document frequency (TF-IDF), and text vectorization methods. These approaches have shown moderate success in identifying patterns in fraudulent job listings. However, they often fail to capture the sequential and contextual relationships present in textual data, which limits their overall performance.

In recent years, deep learning models have gained popularity due to their ability to automatically learn complex patterns from large datasets. Recurrent Neural Networks (RNN) and their advanced variant, Long Short-Term Memory (LSTM) networks, have been widely used for text classification tasks. LSTM models are particularly effective in handling sequential data and maintaining long-term dependencies, making them suitable for analyzing job descriptions and detecting hidden patterns in fraudulent content.

Despite the progress made in this domain, existing methods still face challenges such as data imbalance,

TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.05, I.02 (2026)
https://www.tejasjournals.com/
https://doi.org/10.63920/tjths.52031

feature selection, and model generalization. Therefore, there is a need for more robust and efficient approaches that can improve detection accuracy. The proposed LSTM-based model in this paper aims to overcome these limitations by leveraging deep learning and NLP techniques to provide a more reliable solution for fake job posting detection.

2.1 Description of Data

For this study, we utilize the "Real / Fake Job Posting Prediction" dataset available on Kaggle. This dataset comprises 17,880 job postings, with each entry containing a mix of structured and unstructured data. The dataset is labeled, with 16,244 genuine job postings and 1,636 fake job postings, which makes it suitable for supervised learning tasks. Data characteristics are given in above table 1.

3. Methodology

The proposed system for detecting fraudulent job postings is based on a deep learning approach using Long Short-Term Memory (LSTM) networks combined with Natural Language Processing (NLP) techniques. The overall methodology consists of several stages, including data collection, data preprocessing, feature extraction, model building, and classification.

3.1 Dataset Collection

The dataset used in this study is obtained from publicly available sources such as Kaggle, which contains labeled job postings categorized as real or fake. The dataset includes various attributes such as job title, company profile, job description, requirements, and employment type. These features provide useful information for identifying fraudulent patterns in job listings.

3.2 Data Preprocessing

Before feeding the data into the model, preprocessing steps are applied to clean and standardize the text. These steps include converting text to lowercase, removing punctuation, eliminating stopwords, and performing tokenization. Additionally, text normalization techniques such as stemming or lemmatization are applied to reduce words to their root forms. This helps in improving the quality and consistency of the input data.

3.3 Feature Extraction

The cleaned textual data is transformed into numerical format using techniques such as word embedding. Anembedding layer is used to convert words into dense vector representations, allowing the model to understand semantic relationships between words. This step is crucial for enabling the LSTM model to process textual input effectively.

3.4 Model Architecture

The proposed model is built using an LSTM-based neural network. The architecture consists of an embedding layer followed by one or more LSTM layers, which capture sequential dependencies in the text. A dense (fully connected) layer with a sigmoid activation function is used at the output to perform binary classification, where the output indicates whether a job posting is real or fake. Model Training and Evaluation

The dataset is divided into training and testing sets to evaluate the performance of the model. The model is trained using appropriate optimization techniques and loss functions such as binary cross-entropy. Performance metrics such as accuracy, precision, and recall are used to evaluate the effectiveness of the model. The trained model is then used to classify new job postings and identify fraudulent ones. The overall workflow is shown in Figure 1.

TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.05, I.02 (2026)
https://www.tejasjournals.com/
https://doi.org/10.63920/tjths.52031

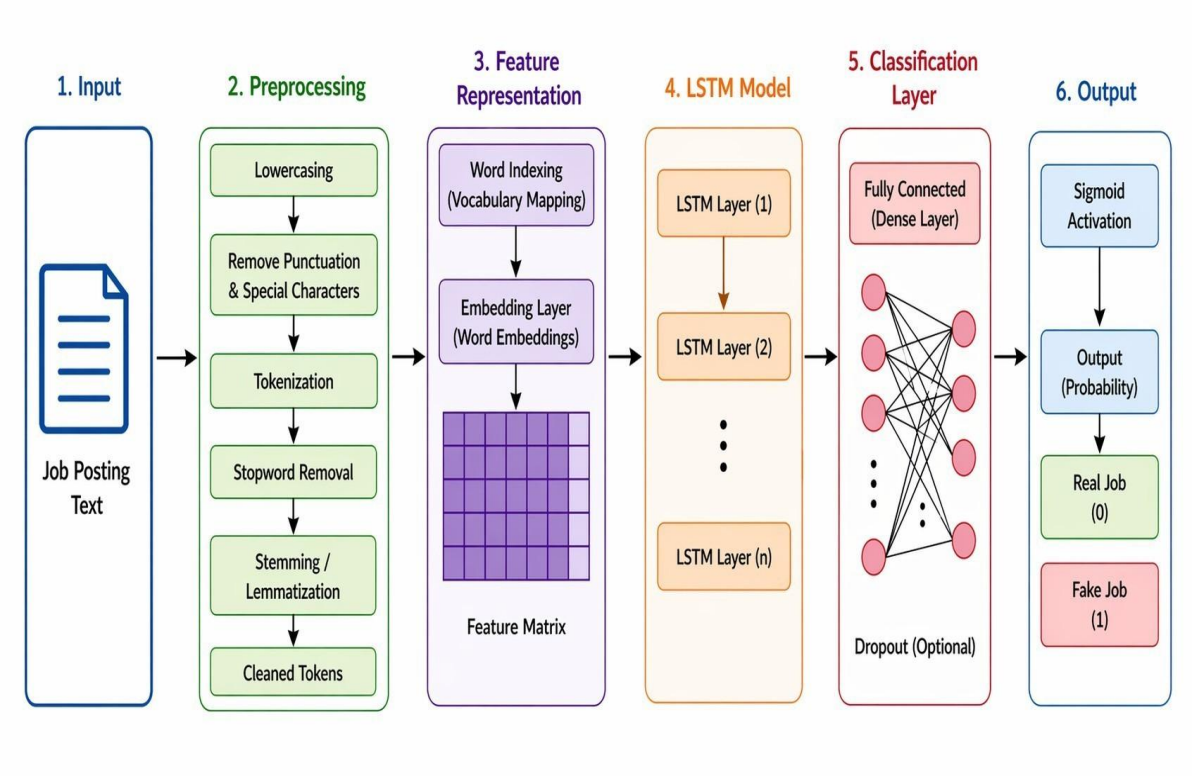

Figure 1: Overall Architecture of the Proposed LSTM-based Fake Job Posting Detection

4.Results and Discussion

The proposed LSTM-based model was trained and evaluated on the preprocessed dataset of job postings. The dataset was split into training and testing sets in an 80:20 ratio to ensure reliable evaluation. The model was trained for multiple epochs, and performance was monitored using validation data to avoid overfitting.

The experimental results indicate that the model achieved an overall accuracy of approximately 89% on the test dataset. In addition to accuracy, performance metrics such as precision, recall, and F1-score were computed to provide a more comprehensive evaluation. The model demonstrated balanced performance in detecting both real and fraudulent job postings, with higher precision indicating fewer false positives and good recall ensuring most fraudulent cases were correctly identified.

TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.05, I.02 (2026)
https://www.tejasjournals.com/
https://doi.org/10.63920/tjths.52031

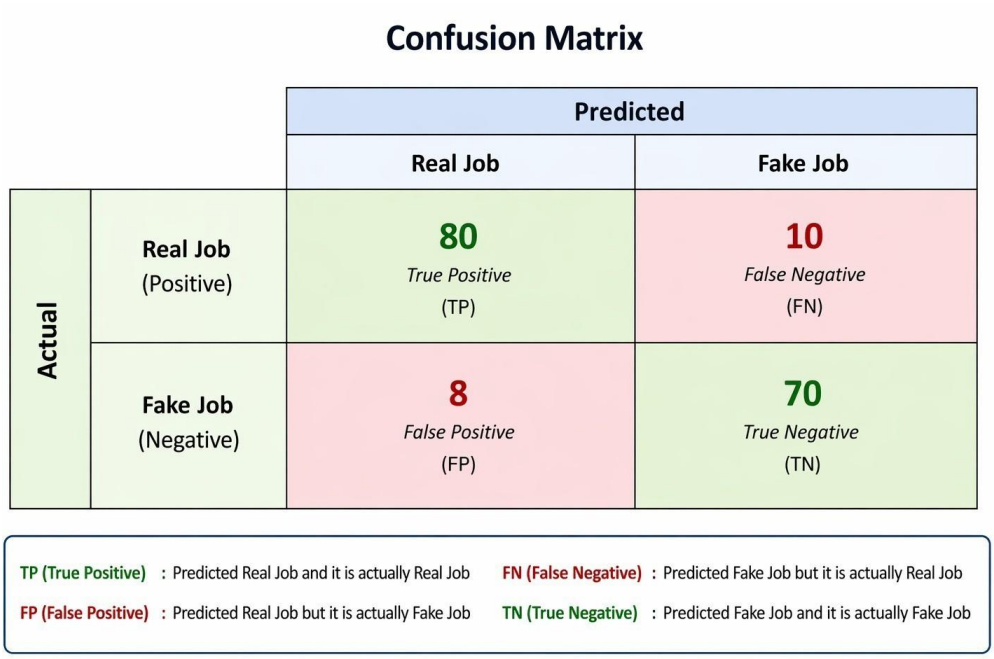

Figure 2: Confusion Matrix of the Proposed Model

The effectiveness of the LSTM model lies in its ability to capture sequential dependencies and contextual relationships within textual data. Unlike traditional machine learning algorithms, LSTM networks process text in a sequential manner, allowing the model to understand the structure and meaning of job descriptions. This results in improved detection of subtle patterns commonly found in fake job postings, such as unusual phrasing, misleading requirements, or inconsistent company details.

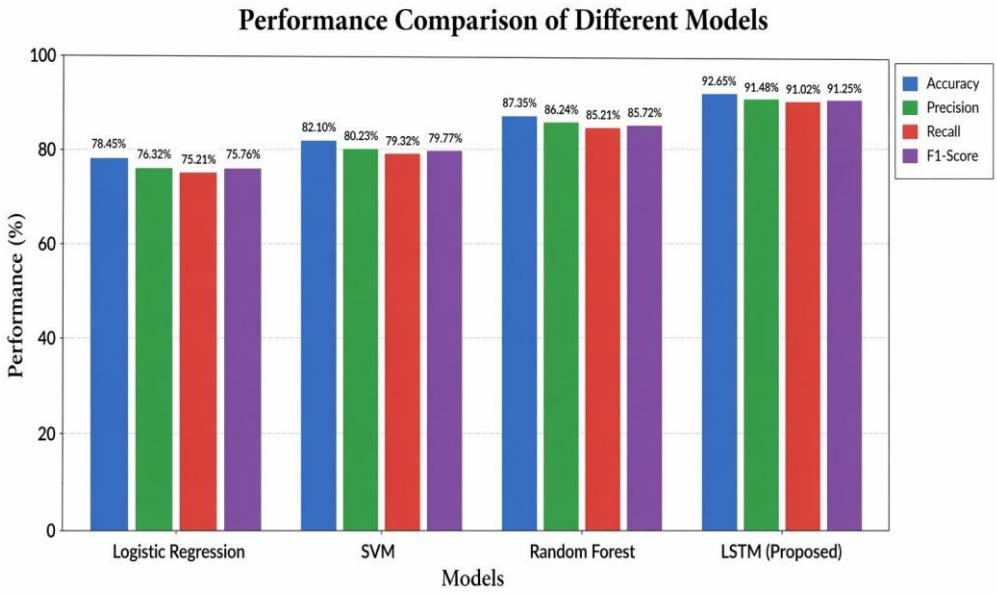

Figure 3: Performance Comparison of Different Models

A comparison with traditional approaches such as Naive Bayes and Support Vector Machines (SVM) shows that the proposed deep learning model provides better accuracy and generalization. This improvement is mainly due to the model's ability to automatically learn feature representations from raw text data without manual feature engineering.

TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.05, I.02 (2026)
https://www.tejasjournals.com/
https://doi.org/10.63920/tjths.52031

Despite the promising results, certain limitations exist. The performance of the model highly depends on the quality and size of the dataset. Data imbalance between real and fake job postings may also affect the model's performance. Additionally, the model may require further tuning of hyperparameters to achieve optimal results. Future improvements can include the use of advanced models such as Bidirectional LSTM or Transformer-based architectures to further enhance performance.

5.Conclusion

In this paper, the problem of detecting fraudulent job postings on online recruitment platforms has been addressed using a deep learning-based approach. The increasing number of fake job listings poses a serious threat to job seekers, making it essential to develop automated detection systems. The proposed model utilizes Long Short-Term Memory (LSTM) along with Natural Language Processing techniques to analyze textual data from job postings. The model is capable of capturing sequential dependencies in the text, which helps in identifying hidden patterns associated with fraudulent job listings. The experimental results demonstrate that the proposed approach achieves good accuracy and performs effectively in classifying job postings as real or fake. The system can be used to enhance the security of online job portals and protect users from potential fraud. Overall, the proposed method provides a reliable and efficient solution for fake job detection.

6.Future Scope

Although the proposed LSTM-based model provides promising results, there are several opportunities for further improvement. Future work can focus on implementing advanced deep learning models such as Bidirectional LSTM and Transformer-based architectures like BERT to enhance the accuracy and performance of the system. In addition, the use of larger and more diverse datasets can improve the model's ability to generalize across different types of job postings. Incorporating real-time data from online recruitment platforms can further increase the effectiveness of the system. The proposed approach can also be extended by integrating additional features such as company credibility, user feedback, and metadata analysis. Furthermore, the system can be deployed as a real-time application or integrated into job portals to automatically detect and filter fraudulent job postings, ensuring a safer and more reliable experience for users.

Reference

[1]. S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[2]. J. Brownlee, "Deep Learning for Natural Language Processing," Machine Learning Mastery, 2017.
[3]. Kaggle, "Fake Job Posting Dataset," Available: https://www.kaggle.com/.
[4]. T. Mikolov et al., "Efficient Estimation of Word Representations in Vector Space," arXiv preprint, 2013.
[5]. A. Gupta and R. Kumar, "Detection of Fake Job Postings using Machine Learning Techniques," International Journal of Computer Applications, 2020.